International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET207446

A Fast-Effective Algorithm on A Concise Representation of Top Rated Utility Mining Datasets

Shaik Subani[1], B. V. N. Krishna Suresh[2]

[1]M.Tech Scholar Department of CSE, NRI Institute of Technology Visadala, Guntur(Dt), Andhra Pradesh, India

[2]Assistant Professor Department of CSE, NRI Institute of Technology Visadala, Guntur(Dt), Andhra Pradesh, India

ABSTRACT

Article Info
Volume 7 Issue 4
Page Number : 188-194
Publication Issue :
July-August-2020

Article History
Accepted : 01 Aug 2020
Published : 05 Aug 2020

The average of customer ratings on a product, which we call a reputation, is one of the key factors in online purchasing decisions. There is, however, no guarantee of the trustworthiness of a reputation since it can be manipulated rather easily. In this paper, we define false reputation as the problem of a reputation being manipulated by unfair ratings and design a general framework that provides trustworthy reputations. For this purpose, we propose Trust-reputation, an algorithm that iteratively adjusts a reputation based on the confidence of customer ratings. We also show the effectiveness of Trust-reputation through extensive experiments in comparisons to state-of-the-art approaches.

Keywords : Battery Storage, Super capacitors, Renewable Resources, Wind Power, Supervisory Controller, Battery Lifetime

I. INTRODUCTION

While using online shopping channels, consumers share their purchasing experiences regarding both goods and services with other potential buyers via evaluation. The most common way for consumers to express their level of satisfaction with their purchases is through online ratings. The overall buyers' satisfaction is quantified as the aggregated score of all ratings and is available to all potential buyers. In this paper, we call this aggregated score for a product its reputation. The reputation of a product plays an important role as a guide for potential buyers and significantly influences consumers' final purchasing decisions. "*Is the Product's Reputation Trustworthy?*" Reputation is the score of a product obtained through collective intelligence, i.e., the result of collaboration between many individuals. The proposed framework, on the other hand, uses all ratings .It evaluates the level of trustworthiness (confidence) of each rating and adjusts the reputation based on the confidence of ratings. We have developed an algorithm that iteratively adjusts a reputation based on the confidence of customer ratings. By adjusting a reputation based on the confidence scores of all ratings, the proposed algorithm calculates the reputation without the risk of omitting ratings by normal users while reducing the influence of unfair ratings by abusers. We call this algorithm, which solves the false reputation problem by computing the true reputation, TRUE-REPUTATION. The computation of a trustworthy reputation starts by measuring the confidence of a rating. We have surveyed previous social science studies that analyzed the characteristics of reliable online information and adopted three key characteristics that are suitable for determining the confidence of a rating [6], [23]. To determine the confidence of a rating, therefore, we have adopted three key factors of activity, objectivity, and consistency and defined these factors in the context of online ratings. First, the user who rates more items displays a higher level of activity. The above description of activity implies that the activity is defined by the amount of interactions between an information producer and the users obtaining his information. There exist, however, no interactions between users in an online rating system; instead, there are actions by users on products. Therefore, we measure user activity in an online rating system based on the amount of actions by the user on products (i.e., the number of products he rates). The objectivity of a rating is calculated based on the deviation of the "rating" from the "reputation" of the product. The difficulty in computing a reputation lies in the fact that the reputation itself is the sum of the ratings adjusted by the confidence, and the confidence of an individual rating is computed using the objectivity of the rating, which uses the reputation in its computation. In other words, the reputation and the confidence of a rating interact with each other in mutual reinforcement. We propose TRUE-REPUTATION, an iterative method, to compute these measures. The contributions of this paper are as follows. First, we have defined false reputation and categorized various real-life scenarios in which a false reputation can occur. The categorization of the false-reputation scenarios helps us design experimental scenarios similar to real-life situations. Second, we have proposed a general framework to address a false reputation by quantifying the level of confidence of a rating. The framework includes TRUE-REPUTATION, an algorithm that iteratively adjusts the reputation based on the confidence of customer ratings. Third, we have verified the superiority of TRUE-REPUTATION by comparing it with machine-learning based algorithms through extensive experiments. Despite their many advantages, e-Businesses lag behind brick and mortar businesses in several fundamental respects. This paper concerns one of these: relationships based on trust and reputation. Recent studies on simple reputation systems for e-Businesses such as eBay have pointed to the importance of such rating systems for deterring moral hazard and encouraging trusting interactions. However, despite numerous studies on trust and reputation systems, few have taken studies across disciplines to provide an integrated account of these concepts and their relationships. This paper first surveys existing literatures on trust, reputation and a related concept: reciprocity. Based on sociological and biological understandings of these concepts, a computational model is proposed. This model can be implemented in a real system to consistently calculate agents' trust and reputation scores.

II. RELATED WORK

Numerous studies have been conducted to improve the trustworthiness of online shopping malls by detecting abusers who have participated in the rating system for the sole purpose of manipulating the information provided to potential buyers (e.g., reputations of sellers and recommended items). Especially in the fields of multi agent and recommendation systems, various strategies have been proposed to handle abusers who attack the vulnerability of the system. Multi agent systems compute and publish the reputation scores of sellers based on a collection of buyer opinions (which can be viewed as ratings). Strategies for improving the robustness of multi agent systems can be classified into two categories. The first group of strategies is based on the principle of majority rule. Considering the collection of majority opinions (more than half the opinions) as fair, this group of strategies excludes the collection of minority opinions, viewed as biased, when calculating the reputation [2], [24], [29]. Despite the obvious usefulness of trust and reputation, conceptual gaps exist in current models about them. Resnick and Zeckhauser (2000b) have pointed out the so called *Pollyanna* effect in their study of the eBay reputation reporting system. This effect refers to the disproportionately positive feedbacks from users and rare negative feedbacks. They have also pointed out that despite the incentives to free ride (for not providing feedbacks), feedbacks by agents are provided in more than half of the transactions.

III. UNDERSTANDING TRUST AND REPUTATION

Trust and reputation have become important topics of research in many fields. This section reviews a few of the important studies. Sciento metrics refers to the study of measuring research outputs such as journal impact factors. Reputation as used by this community usually refers to number of cross citations that a given author or journal has accumulated over a period of time (Garfield, 1955). As pointed out by Makino, *et al.*, 1998 and others, cross citation is a reasonable but sometimes confounded measure of one's reputation. Economists have studied reputation in game theoretic settings. Entry deterrence is one of the early areas for game theorists' study of reputation. Kreps and Wilson (1982) postulate that imperfect information about players' payoffs creates "reputation effects" for multi-stage games. They claim that an incumbent firm seeks to acquire an early reputation for being "tough" in order to decrease the probability for future entries into the industry. Milgrom an d Roberts (1982) report similar findings by using asymmetric information to explain the reputation phenomenon. For an incumbent firm, it is rational to seek a "predation" strategy for early entrants even if "it is costly when viewed in isolation, because it yields a reputation which deters other entrants." (*ibid.*). Whether online reputation systems contribute to trade is answered by several research analysis of existing systems. Resnick and Zeckhauser (2000b) have analyzed the feedback rating used in eBay as a reputation system. "Reputation" is taken to be a function of the cumulative positive and non-positive ratings for a seller or buyer. Trust by one agent of another is inferred by an implicit mechanism. They have found that the system does encourage transactions. Houser and Wooders (2000) have studied auctions in eBay and describe reputation as the *propensities to default* – for a buyer, it is the probability that if the buyer wins, he will deliver the payment as promised before the close of the auction; for a seller, it is the probability that once payment is received, he will deliver the item auctioned. Their economic analysis shows that reputation has a statistically significant effect on price. Unfortunately, they did not model how reputation is built; nor how trust is derived from reputation. Both Lucking-Reily, *et al.* (1999) and Bajari and Hortacsu (2000) have examined coin auctions in eBay. These economic studies have provided empirical confirmation of reputation effects in internet auctions. Bajari and Hortacsu (2000) have also reported the "winner's curse" phenomenon in their analysis. This phenomenon refers to a fall in the bidder's expected profits when the expected number of bidders is increased. Among sociologists, reputation as a quantitative concept is often studied as a network parameter associated with a society of agents (Wasserman and Faust, 1994). Reputation or prestige is often measured by various centrality measures. An example is a measure proposed by Katz (1953) based on a stochastic coincidence matrix where entries record social linkages among agents. Because the matrix is stochastic, the right eigenvector associated with the eigenvalue of 1 is the stationary distribution associated with the stochastic matrix (Strang, 1988). The values in the eigenvector represent the reputations of the individuals in the society.Unfortunately, these values are often global in nature, and lacks context dependence.

IV. MODEL RATIONAL

There are many reciprocity strategies proposed by game-theoreticians; the most famous of which is the tit-for-tat strategy which has been extensively studied in the context of the Prisoners's Dilemma game (Axelrod, 1984; Pollock and Dugatkin, 1992; Nowak and Sigmund, 2000). Not everyone in a society learns the same norms in all situations. Structural variables affect individuals' level of confidence and willingness to reciprocate. In the case of cooperation, some cooperate only in contexts where they expect reciprocation from their interacting parties. Others will only do so when they are publicly committed to an agreement.

When facing social dilemmas[i], trustworthy individuals tend to **trust** others with a reputation for being trustworthy and shun those deemed less so (Cosmides and Tooby, 1992). In an environment where individuals "regularly" perform **reciprocity** norms, there is an incentive to acquire a **reputation** for reciprocative actions (Kreps, 1990; Milgrom, *et al.*, 1990; Ostrom, 1998). "Regularly" refers to a *caveat* observed by sociologists that reputation only serves a normative function in improving the fitness of those who cooperate while disciplining those who defect if the environment encourages the spreading of reputation information (Castelfranchi, *et al.*, 1998). In the words of evolutionary biologists, having a good reputation increases an agent's *fitness* in an environment where reciprocity norms are expected (Nowak and Sigmund, 1998). Therefore, developing the quality for being trustworthy is an asset since trust affects how willing individuals are to participate in reci procative interactions (Dasgupta, 2000; Tadelis, 1999).

This paper uses the following definition for reciprocity:

Reciprocity: mutual exchange of deeds (such as favor or revenge).This definition is largely motivated by the many studies of reciprocity in which repeated games are played between two or more individuals (Raub and Weesie, 1990; Boyd and Richersen 1989; Nowak and Sigmund, 1998). Two types of reciprocity are considered: direct reciprocity refers to interchange between two concerned agents; indirect reciprocity refers to interchange between two concerned agents interceded by mediating agents in between. Reciprocity can be measured in two ways. Firstly, reciprocity can be viewed as a social norm shared by agents in a society. The higher this "societal reciprocity," the more likely one expects a randomly selected agent from that society to engage in reciprocating actions. Secondly, reciprocity can be viewed as a dyadic variable between two agents (say *ai* and *aj*). The higher this "dyadic reciprocity," the more one expects *ai* and *aj* to reciprocate each other's actions. In this latter case, no expectation about other agents should be conveyed. For any single agent *ai*, the cumulative dyadic reciprocity that *ai* engages in with other agents in a society should have an influence on *ai*'s *reputation* as a reciprocating agent in that society.

Reputation: Perception that an agent creates through past actions about its intentions and norms [ii].Reputation is a social quantity calculated based on actions by a given agent *ai* and observations made by others in an "embedded social network" that *ai* resides (Granovetter, 1985). *ai*'s reputation clearly affects the amount of trust that others have toward it. How is trust defined? The definition for trust by Gambetta (1988) is often quoted in the literature: "… trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [it] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [the agent's] own action" (*ibid.*). This paper elects the term "subjective expectation" rather than "subjective probability" to emphasize the point that trust is a summary quantity that an agent has toward another based on *a number of former encounters* between them:

Trust: A subjective expectation an agent has about another's future behavior based on the history of their encounters.Trust is a subjective quantity calculated based on the two agents concerned in a dyadic encounter. Dasgupta (2000) gave a similar definition for trust: the expectation of one person about the actions of others that affects the first person's choice, when an action must be taken before the actions of others are known.

V. RELIABLE AND EFFICIENT MECHANISM FOR REPUTATION

The last section has considered how reputation can be determined when two agents are concerned. This section extends the analysis to arbitrary number of agents. A schematic diagram of an embedded social network for agents *a* and *b*.

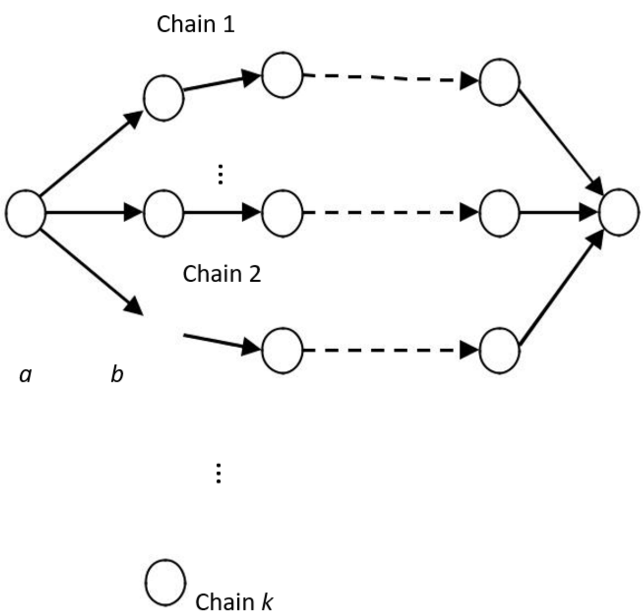

Figure 2. Illustration of a parallel network between two agents a and b.

Figure 2 shows a parallel network of k chains between two agents of interest, where each chain consists of at least one link. Agent a would like to estimate agent b's reputation as defined by the embedded network between them.[iii] Clearly, to combine the parallel evidence about b, measures of "reliability" are required to weight all the evidences. For each chain in the parallel network, how should the total weight be tallied? Two possible methods are plausible: additive and multiplicative. The problem with additive weight is that if the chain is "broken" by a highly unreliable link, the effect of that unreliability is local to the immediate agents around it. In a long social chain however, an unreliability chain is certain to cast serious doubt on the reliability of any estimate taken from the chain as a whole. On the other hand, a multiplicative weighting has "long-distance" effect in that an unreliable link affects any estimate based on a path crossing that link.

VI. CONCLUSION

This paper has surveyed the literatures on trust and reputation models across diverse disciplines. A number of significant shortcomings of these models have been pointed out. We have attempted to integrate our understanding across the surveyed literatures to construct a computational model of trust and reputation.

Our model has the following characteristics:

- makes explicit the difference between trust and reputation
- defines reputation as a quantity relative to the particular embedded social network of the evaluating agent and encounter history
- defines trust as a dyadic quantity between the trust or and the trustee which can be inferred from reputation data about the trustee
- proposes a probabilistic mechanism for inference among trust, reputation, and level of reciprocity

The explicit formulation of trust, reputation, and related quantities suggests a straightforward implementation of the model in a multi-agent environment (such as an electronic market).Two immediate future works follow what is presented in this paper. Firstly, the propagation mechanism for reputation only applies to parallel networks. Extending the mechanism to arbitrary graphs with reasonable computational complexity would generalize the model proposed here. A forthcoming paper addresses this mechanism. Secondly, although context is explicitly modeled in the parameters studied here, cross-contexts estimation for the parameters in our model is not addressed. A simple scheme is to create vectorized versions of the quantities studied in this paper. More complex schemes would involve semantic inferences across different contexts.

VII.REFERENCES

[1]. A. Abdul-Rahman, S. Hailes (2000) “Supporting Trust in Virtual Communities,” Proc. 33rd Hawaii International Conference on System Sciences.

[2]. G. Akerlof (1970) “The Market for ‘Lemons’: Qualitative Uncertainty and the Market Mechanism,” Quarterly Journal of Economics, 84, pp. 488-500.

[3]. R. Axelrod (1984) The Evolution of Cooperation. New York: Basic Books.

[4]. P. Bajari, A. Hortacsu (1999) “Winner’s Curse, Reserve Prices and Endogenous entry: Empirical Insights from eBay Auctions,” Stanford Institute for aEconomic Policy Research (SIEPR) Policy paper No. 99-23.

[5]. J. H. Barkow, L. Cosmides, J. Tooby (eds.) (1992) The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford: Oxford University Press

[6]. L. C. Becker, (1990) Reciprocity. Chicago: University of Chicago Press.

[7]. K. Binmore (1997) "Rationality and Backward Induction," Journal of Economic Methodology, 4, pp. 23 -41.

[8]. R. Boyd and P. J. Richerson (1989) “The Evolution of Indirect Reciprocity,” Social Networks, 11, pp. 213-236.

[9]. C. Castelfranchi, R. Conte, M. Paolucci (1998) “Normative Reputation and the Costs of Compliance,” Journal of Artificial Soci eties and Social Simulations, 1(3).

[10]. S. Krishnamoorthy, “Pruning strategies for mining high utility itemsets,” Expert Syst. Appl., vol. 42, no. 5, pp. 2371-2381, 2015.

[11]. C. Lin, T. Hong, G. Lan, J. Wong, and W. Lin, “Efficient updating of discovered high-utility itemsets for transaction deletion in dynamic databases,” Adv. Eng. Informat., vol. 29, no. 1, pp. 16-27, 2015.

[12]. G. Lan, T. Hong, V. S. Tseng, and S. Wang, “Applying the maximum utility measure in high utility sequential pattern mining,” Expert Syst. Appl., vol. 41, no. 11, pp. 5071-5081, 2014.

[13]. Y. Liu, W. Liao, and A. Choudhary, “A fast high utility itemsets mining algorithm,” in Proc. Utility-Based Data Mining Workshop, 2005, pp. 90-99.

[14]. M. Liu and J. Qu, “Mining high utility itemsets without candidate generation,” in Proc. ACM Int. Conf. Inf. Knowl. Manag., 2012, pp. 55-64.

[15]. J. Liu, K. Wang, and B. Fung, “Direct discovery of high utility itemsets without candidate generation,” in Proc. IEEE Int. Conf. Data Mining, 2012, pp. 984-989.

[16]. Y. Lin, C. Wu, and V. S. Tseng, “Mining high utility itemsets in big data,” in Proc. Int. Conf. Pacific-Asia Conf. Knowl. Discovery Data Mining, 2015, pp. 649-661.

[17]. Y. Li, J. Yeh, and C. Chang, “Isolated items discarding strategy for discovering high-utility itemsets,” Data Knowl. Eng., vol. 64, no. 1, pp. 198-217, 2008.

[18]. J. Pisharath, Y. Liu, B. Ozisikyilmaz, R. Narayanan, W. K. Liao, A. Choudhary, and G.

Memik, NU-MineBench version 2.0 dataset and technical report Online]. Available: http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html, 2005.

[19]. G. Pyun and U. Yun, “Mining top-k frequent patterns with combination reducing techniques,” Appl. Intell., vol. 41, no. 1, pp. 76-98, 2014.

[20]. T. Quang, S. Oyanagi, and K. Yamazaki, “ExMiner: An efficient algorithm for mining top-k frequent patterns,” in Proc. Int. Conf. Adv. Data Mining Appl., 2006, pp. 436 - 447.

[21]. H. Ryang and U. Yun, “Top-k high utility pattern mining with effective threshold raising Strategies,” Knowl.-Based Syst., vol. 76, pp. 109-126, 2015.

[22]. H. Ryang, U. Yun, and K. Ryu, “Discovering high utility itemsets with multiple minimum supports,” Intell. Data Anal., vol. 18, no. 6, pp. 1027-1047, 2014.

[23]. B. Shie, H. Hsiao, V. S. Tseng, and P. S. Yu, “Mining high utility mobile sequential patterns in mobile commerce environments,” in Proc. Int. Conf. Database Syst. Adv. Appl. Lecture Notes Comput. Sci., 2011, vol. 6587, pp. 224-238.

[24]. P. Tzvetkov, X. Yan, and J. Han, “TSP: Mining top-k closed sequential patterns,” Knowl. Inf. Syst., vol. 7, no. 4, pp. 438-457, 2005.

[25]. V. S. Tseng, C. Wu, B. Shie, and P. S. Yu, “UP-Growth: An efficient algorithm for high utility itemset mining,” in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 253-262.

[26]. V. S. Tseng, C. Wu, P. Fournier-Viger, and P. S. Yu, “Efficient algorithms for mining the concise and lossless representation of high utility itemsets,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 3, pp. 726-739, Mar. 1, 2015.

[27]. C. Wu, P. Fournier-Viger, P. S. Yu, and V. S. Tseng, “Efficient mining of a concise and lossless representation of high utility itemsets,” in Proc. IEEE Int. Conf. Data Mining, 2011, pp. 824-833.

[28]. J. Wang and J. Han, “TFP: An efficient algorithm for mining top-k frequent closed itemsets,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 5, pp. 652-663, May 2005.

[29]. C. Wu, Y. Lin, P. S. Yu, and V. S. Tseng, “Mining high utility episodes in complex event sequences,” in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 536-544.

[30]. C. Wu, B. Shie, V. S. Tseng, and P. S. Yu, “Mining top-k high utility itemsets,” in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 78-86.

Cite this article as :

Shaik Subani, B. V. N. Krishna Suresh, " A Fast-Effective Algorithm on A Concise Representation of Top Rated Utility Mining Datasets, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 7, Issue 4, pp.188-194, July-August-2020. Available at doi : https://doi.org/10.32628/IJSRSET207446
Journal URL : http://ijsrset.com/IJSRSET207446